

WHAT IS CLAIMED:

1. A linguistic segmentation tool comprising:
 - a lexical feature extraction component configured to receive text and generate lexical feature vectors relating to the text, the lexical feature vectors including words from the text and syntactic classes of the words;
 - an acoustic feature extraction component configured to receive an audio version of the text and generate acoustic feature vectors relating to the audio version of the text; and
 - a statistical framework component configured to generate linguistic features associated with the text based on the acoustic feature vectors and the lexical feature vectors.
2. The linguistic segmentation tool of claim 1, wherein the linguistic features include periods, quotation marks, exclamation marks, commas, and phrasal boundaries.
3. The linguistic segmentation tool of claim 1, further comprising:
 - a transcription component configured to generate the text based on the audio version of the text.
4. The linguistic segmentation tool of claim 1, wherein the statistical framework includes:

an acoustic model configured to estimate a probability of an occurrence of the linguistic features based on the acoustic feature vectors.

5. The linguistic segmentation tool of claim 4, wherein the statistical framework includes:

a language model configured to estimate a probability that one of the lexical feature vectors corresponds to a text boundary.

6. The linguistic segmentation tool of claim 5, wherein the statistical framework includes:

a maximum likelihood estimator configured to generate the linguistic features based on the probabilities generated by the acoustic model and the language model.

7. The linguistic segmentation tool of claim 1, wherein the lexical feature vectors additionally include an identification of a structured speech member of the word.

8. The linguistic segmentation tool of claim 1, wherein the acoustic feature vectors are based on prosodic features including at least one of pause, rate, energy, and pitch.

9. The linguistic segmentation tool of claim 1, wherein the syntactic classes are indicative of a role of the word in the text.

10. The linguistic segmentation tool of claim 9, wherein the syntactic classes include syntactic classes based on affixes of the words.

11. The linguistic segmentation tool of claim 10, wherein the syntactic classes include syntactic classes based on frequently occurring words.

12. A method for determining linguistic information for words corresponding to a transcribed version of an audio input stream including speech, the method comprising:

generating lexical features for the words, including a syntactic class associated with at least one of the words;

generating acoustic features for the audio input stream, the acoustic features being based on at least one of speaker pauses, speaker rate, speaker energy, and speaker pitch; and

generating the linguistic information based on the lexical features and the acoustic features.

13. The method of claim 12, further comprising:
automatically transcribing the audio input stream to generate the words corresponding to the transcribed version of the speech.

14. The method of claim 12, further comprising:
creating a language model configured to estimate a probability that the lexical features correspond to a word boundary based on the lexical features.
15. The method of claim 14, further comprising:
creating an acoustic model configured to estimate a probability of an occurrence of the linguistic information based on the acoustic features.
16. The method of claim 15, wherein generating the linguistic information based on the lexical features and the acoustic features includes using a maximum likelihood estimator configured to estimate a final probability of an occurrence of the linguistic information based on the probabilities generated by the acoustic model and the language model.
17. The method of claim 12, wherein the syntactic class is indicative of the role of the at least one of the words.
18. The method of claim 12, wherein the syntactic class is based on affixes of the words.
19. The method of claim 12, wherein the syntactic class is based on word frequency.

20. The method of claim 12, wherein the linguistic information includes periods, quotation marks, exclamation marks, commas, and phrasal boundaries.

21. A computing device for determining linguistic information for words corresponding to a transcribed version of an audio input stream that includes speech, the computing device comprising:

a processor; and

a computer memory coupled to the processor and containing programming instructions that when executed by the processor cause the processor to:

generate lexical features for the words, including a syntactic class associated with at least one of the words,

generate acoustic features for the audio input stream, the acoustic features being based on at least one of speaker pauses, speaker rate, speaker energy, and speaker pitch,

generate the linguistic information based on the lexical features and the acoustic features, and

output the generated linguistic information as meta-information embedded in the transcribed version of the audio input stream.

22. The computing device of claim 21, wherein the syntactic class is indicative of the role of the at least one of the words.

23. The computing device of claim 21, wherein the syntactic class is based on affixes of the words.

24. The computing device of claim 21, wherein the syntactic class is based on word frequency.

25. A method for associating meta-information with a document transcribed from speech, the method comprising:

building a language model based on lexical feature vectors extracted from the document, the lexical feature vectors including words and syntactic classifications of the words;

building an acoustic model based on acoustic feature vectors extracted from the speech; and

combining outputs of the language model and the acoustic model in a statistical framework that estimates a probability for associating the meta-information with the document.

26. The method of claim 25, wherein the meta-information relates to linguistic features of the document.

27. The method of claim 26, wherein the linguistic features include periods, quotation marks, exclamation marks, commas, and phrasal boundaries.

28. The method of claim 25, wherein the acoustic feature vectors are based on prosodic features including pause, rate, energy, and pitch.

29. The method of claim 25, wherein the syntactic class is indicative of the role of the at least one of the words.

30. The method of claim 25, wherein the syntactic class is based on affixes of the words.

31. The method of claim 25, wherein the syntactic class is based on word frequency.

32. A device comprising:

means for building a language model based on lexical feature vectors extracted from a document transcribed from human speech, the lexical feature vectors including a word and a syntactic classification of the word;

means for building an acoustic model based on acoustic feature vectors extracted from the speech; and

means for combining outputs of the language model and the acoustic model to estimate a probability for associating a linguistic feature with the document.

33. A computer-readable medium containing program instructions for execution by a processor, the program instructions, when executed by the processor, cause the processor to perform a method comprising:

generating lexical features for words corresponding to a transcribed version of speech, the lexical features including a syntactic class associated with at least one of the words;

generating acoustic features for the speech, the acoustic features based on at least one of speaker pauses, speaker rate, speaker energy, and speaker pitch; and

generating linguistic information for the words based on the lexical features and the acoustic features.